Latent instrumental variables

Ebbes, P.

Publication date:
2004

Chapter 5

Estimating the return to education using LIV

5.1 Introduction

We apply the methods developed in the previous chapters to three empirical datasets to examine the return to education on income. Education is an important topic in public debates. Over the past decades, much research has been conducted to estimate the causal effect of education on earnings, see for instance Griliches (1977), Card (1999, 2001), or Uusitalo (1999). Most of the studies in question have focused on estimating a version of the following linear regression equation:

$$y_i \quad = \quad \beta_0 + \beta_1 S_i + X_i \beta_2 + \epsilon_i, \tag{5.1}$$

where y_i is the logarithm of a measure of earnings, S_i is a measure of education and X_i is a collection of other explanatory variables assumed to influence y_i. β_1 measures the effect of education on income and is expected to be positive. The disturbances ϵ_i represent all other influences not explicitly accounted for. If the disturbances are distributed independently of the explanatory variables S_i and X_i, the simple OLS estimator can be used to estimate β_1. However, the independence assumption may not be realistic and in this case it can be shown that the OLS estimator is biased (e.g. Greene, 2000). Four major potential sources of bias (ability bias, measurement error bias, heterogeneity bias, and optimizing behavior bias) have been identified in the literature on the relationship between education and income, each of which is discussed in section 5.2. As will become clear from this discussion, there is little agreement on the direction and magnitude of the potential bias in the OLS estimator of the return to education effect. This situation is not surprising in view of the many sources of potential regressor-error dependencies, with each of them having their own specific impact on the direction and magnitude of the bias in OLS. A further complicating factor is that these causes offset or enforce each other.

One way to circumvent problems of endogeneity is to find instruments and apply two-stage least squares or limited maximum likelihood estimation techniques (see e.g. Bowden and Turkington, 1984, Verbeek, 2000, or Greene, 2000). Instruments are variables that mimic the troublesome regressors as well as possible but are uncorrelated with the error term. Hence, instrumental variables cannot have a direct effect on the dependent variable. In practice it is not obvious how or where to find valid instruments. Furthermore, instruments are often weak, i.e. they only explain a small part of the variance of the endogenous regressor. This may result in estimates that are even more biased than the OLS estimates (Staiger and Stock, 1997, Bound, Jaeger and Baker, 1995), see also sections 4.3 and 4.4. In section 5.3 we discuss the problems of IV estimation for the model given in (5.1), and it is shown that instrumental variable estimation in estimating the return to education is not a straightforward exercise.

Card (1999, 2001) surveys several empirical studies on the return to education and finds regression estimates ranging from about $0.03 - 0.14$. Quite often, the OLS estimates were not found to be statistically different from the instrumental variable estimates. As suggested above and discussed in more detail in section 5.3, instrumental variables estimates for these kind of studies are potentially biased as well, because the instruments used are possible weak and endogenous. We find empirical evidence for this in section 5.4. Recent evidence from twin studies suggests an upward bias in the OLS estimator of about

10%-15% (cf. Card, 1999). The major advantage of using data on twins is that no observed instruments are required as the within–family estimator can be used (see also subsection 5.3.3).

The latent instrumental variable (LIV) method proposed in the previous chapters provides approximately unbiased estimates of the model parameters without relying on observed instruments. It is based on the assumption that a discrete latent variable splits the endogenous regressor x into an exogenous component and an endogenous component that is correlated with the error term. In section 5.4 we estimate the return to education for three datasets using LIV. We show, by using the previously proposed diagnostics, that the LIV model is not particularly sensitive to outliers and fits the data sets fairly well. The instruments that are estimated by LIV from the data are shown to be much stronger than the available observed instruments that are typically used in these applications. Furthermore, we investigate the validity of the available observed instrumental variables in these datasets using the methods proposed in section 4.3. We find considerable evidence that in two of the three applications the observed instruments are weak and/or endogenous. Overall, the LIV approach yields results that are more consistent than the classical IV results. We find a moderate upward bias in OLS of $\approx 7\%$ which is close to recent results from twin studies, and supports the ability bias hypothesis. On the other hand, the bias in OLS, as indicated by the classical IV estimates, ranges from -80% to $+30\%$ for the three applications, which illustrates that opposite answers may be obtained, if one uses different sets of instrumental variables to address the same substantive research question. Section 5.5 presents a summary of our findings.

5.2 Sources of bias in the OLS estimate of the return to education

5.2.1 Ability bias

Much work has focused on the issue whether the presence of a –so called– ‘ability bias’ overstates the true causal effect of education on earnings (e.g.

Angrist and Krueger, 1991, Harmon and Walker, 1995, Verbeek, 2000). 'Ability' can be seen as an omitted variable that enables (certain) individuals to obtain more income. In this case, the true model is $y_i = \beta_0 + \beta_1 S_i + \rho A_i + \epsilon_i$, where A_i denotes 'ability' and ρ is the effect of 'ability' on income which is expected to be positive. We assume for the moment that there are no other explanatory variables present. The probability limit of the OLS estimator for β_1, while omitting 'ability', is[1]

$$\text{plim}\ \hat{\beta}_1^{\text{OLS}} = \text{plim}\ \frac{\sum_{i=1}^{n}(S_i - \bar{S})(y_i - \bar{y})}{\sum_{i=1}^{n}(S_i - \bar{S})^2} = \beta_1 + \rho\frac{\sigma_{SA}}{\sigma_S^2},$$

where σ_{SA} denotes the covariance between education and ability, and σ_S^2 is the variance of S. When individuals with higher ability have chosen to obtain more education ($\sigma_{SA} > 0$), the effect of education on income is overstated, as $\rho > 0$, since the effect of unobserved ability is falsely attributed to it. As such, exogenous shocks in education levels will have less effect on individual wages than what is predicted by the OLS regression model, and education seems more valuable than it actually is.

5.2.2 Measurement error bias

Although 'ability' bias may induce a positive bias in OLS, error in the measurement of the education variable S_i may result in downward biases. Often, the only data available to measure education is 'years of schooling'. However, it can be questioned whether 'years of schooling' adequately measures 'total education'. Griliches (1977) shows that if the measures for education are imperfect, OLS estimates can have a large downward bias. This bias is magnified (even if the error of measurement is small) when more variables are included in the model. This can be seen as follows (Griliches, 1977). Let the true wage-education equation be

$$y_i \quad = \quad \beta_1 S_i^* + X_i \beta_2 + \epsilon_i,$$

[1] Substitute the expression for y_i in the expression for $\hat{\beta}_1^{\text{OLS}}$, complete the terms, and use the law of large numbers and Slutsky's theorems (see Ferguson, 1996).

with $\beta_1 = 0.1$ and $\beta_2 = 0.01$, say. S_i^* is the true but unobserved level of education. The observed level of schooling S_i is measured with error, such that $S_i = S_i^* + u_i$, with u_i a random term independent of S_i^* and ϵ_i. Let $\lambda = \sigma_u^2/\sigma_S^2$ be the fraction of the observed variance in schooling that is due to measurement error and assume that X_i (e.g. ability or other explanatory variables) is measured without error. Regressing y on S, while ignoring X, gives

$$\hat{\beta}_1^{\text{OLS}} = \frac{\sum_{i=1}^{n}(S_i - \bar{S})(y_i - \bar{y})}{\sum_{i=1}^{n}(S_i - \bar{S})^2} = \beta_1 - \beta_1 \frac{\sum_{i=1}^{n}(S_i - \bar{S})(u_i - \bar{u})}{\sum_{i=1}^{n}(S_i - \bar{S})^2} + \\ + \beta_2 \frac{\sum_{i=1}^{n}(S_i - \bar{S})(X_i - \bar{X})}{\sum_{i=1}^{n}(S_i - \bar{S})^2},$$

which has probability limit

$$\text{plim}\ \hat{\beta}_1^{\text{OLS}} \quad = \quad \beta_1 - \beta_1 \frac{\sigma_{SU}}{\sigma_S^2} + \beta_2 \frac{\sigma_{SX}}{\sigma_S^2},$$

where the covariance $\sigma_{SU} = \sigma_U^2$, since S^* is independent of u, and σ_{SX} is the covariance between S and X. Suppose that 10% of the observed variance in schooling is due to measurement error, i.e. $\lambda = 0.1$, that $\sigma_S = 3$, $\sigma_X = 15$, and that the correlation between X and S is $\rho_{XS} = 0.5$. Then the inconsistency of $\hat{\beta}_1^{\text{OLS}}$ follows from

$$\text{plim}\ \hat{\beta}_1^{\text{OLS}} = 0.1 - 0.1 \times 0.1 + 0.01 \times \frac{3 \times 15 \times 0.5}{9} = 0.115,$$

from which it can be seen that the simple OLS estimator is biased upward by 15%. If the additional explanatory variable X is added, the probability limit for $\hat{\beta}_1^{\text{OLS}}$ becomes (see Judge et al., 1985, p.708)

$$\text{plim}\ \hat{\beta}_1^{\text{OLS}} = \beta_1 - \lambda \frac{\beta_1}{1 - R_{SX}^2} = 0.1 - 0.1 \times 0.1/0.75 = 0.087,$$

where R_{SX} is the multiple correlation coefficient between S and X, and OLS exhibits a downward bias of 13%. It can be seen that: (1) measurement error in

S induces a negative bias in the OLS estimator for β_1, (2) this negative bias can be offset by an upward bias due to the omission of X and, hence, the total bias in $\hat{\beta}_1^{\text{OLS}}$ may be positive, (3) adding explanatory variables X correlated with the systematic components of schooling increases the measurement error bias in $\hat{\beta}_1^{\text{OLS}}$, even when the additional variables do not explain much of the variance in the observed (log) wages (Griliches, 1977).

5.2.3 Heterogeneity bias

Heterogeneity in the regression coefficients of (5.1) is a third source of potential bias in the OLS estimates. People differ with respect to their marginal return to education, their marginal cost for education, and their tastes or beliefs. Now the return to education is not a single parameter but a random variable that potentially differs with background characteristics of individuals as well. Unobserved heterogeneity might induce a dependency between S and ϵ. This can be seen as follows, where we omit other explanatory variables and make some simplifying assumption on the distribution of S_i to present a constructive example, see also Card (1999, 2001). Let

$$y_i = \beta_{0i} + \beta_{1i} S_i + \epsilon_i, \tag{5.2}$$

with $\beta_{0i} = \beta_0 + u_{0i}$ and $\beta_{1i} = \beta_1 + u_{1i}$, such that E $(\beta_{0i}) = \beta_0$ and E $(\beta_{1i}) = \beta_1$. Biases in the OLS estimator for schooling arise when the unobserved heterogeneity (u_{0i}, u_{1i}) is correlated with schooling S_i. This can be illustrated as follows. Following Card (1999, 2001), let

$$u_{0i} = \beta_{0i} - \beta_0 = \lambda(S_i - \mu_S) + \nu_{0i}$$
$$u_{1i} = \beta_{1i} - \beta_1 = \psi(S_i - \mu_S) + \nu_{1i},$$

where υ_{0i} and υ_{1i} are independent of each other and of ϵ_i. We also assume that E $(\nu_{0i}|S_i) =$ E $(\nu_{1i}|S_i) = 0$, and $\mu_S =$ E (S_i). Now

$$\lambda = \frac{\text{cov}(\beta_{0i}, S_i)}{\text{var}(S_i)} \quad \text{and} \quad \psi = \frac{\text{cov}(\beta_{1i}, S_i)}{\text{var}(S_i)},$$

and $y_i = \tilde{\beta}_0 + (\beta_1 + \lambda - \psi\mu_S)S_i + \psi S_i^2 + \nu_{0i} + \nu_{1i}S_i + \epsilon_i$, where $\tilde{\beta}_0 = \beta_0 - \lambda\mu_S$. It can be shown that (see also Card, 1999, 2001)[2]

$$\begin{aligned}\hat{\beta}_1^{\text{OLS}} = \frac{\sum_{i=1}^n (S_i - \bar{S})(y_i - \bar{y})}{\sum_{i=1}^n (S_i - \bar{S})^2} = \beta_1 + \lambda - \psi\mu_S + \\ + \psi\frac{\sum_{i=1}^n S_i^3 - n\bar{S}\left(\overline{S \circ S}\right)}{\sum_{i=1}^n (S_i - \bar{S})^2} + \frac{\sum_{i=1}^n \nu_{1i}S_i^2 - n\bar{S}\left(\overline{S \circ \nu_1}\right)}{\sum_{i=1}^n (S_i - \bar{S})^2} + \\ + \frac{\sum_{i=1}^n S_i(\nu_{0i} + \epsilon_i) - n\bar{S}(\bar{\nu} + \bar{\epsilon})}{\sum_{i=1}^n (S_i - \bar{S})^2},\end{aligned}$$

and by using the weak law of large numbers (multiply nominators and denominators by $(1/n)$)[3] and Slutsky's theorem (Ferguson, 1996)

$$\text{plim}\ \hat{\beta}_1^{\text{OLS}} = \beta_1 + \lambda - \psi\mu_S + \psi\,\text{plim}\ \frac{\frac{1}{n}\sum_{i=1}^n S_i^3 - \bar{S}\left(\overline{S \circ S}\right)}{\frac{1}{n}\sum_{i=1}^n (S_i - \bar{S})^2}.$$

The latter fraction is equal to the regression coefficient of S_i^2 on S_i which has probability limit $2\mu_S$ (it is assumed that the distribution of S_i is symmetric)[4]. Hence,

$$\text{plim}\ \hat{\beta}_1^{\text{OLS}} = \beta_1 + \lambda + \psi\mu_S. \tag{5.3}$$

This relation generalizes the conventional analysis of ability bias (Griliches, 1977). If $\beta_{1i} = \beta_1$ for all i, i.e. there is no heterogeneity in the schooling coefficient, then $\psi = 0$ and the inconsistency of the OLS estimate for β_1 is equal to λ, which can be interpreted as the conventional ability bias. If both intercept and slope vary across individuals, i.e. $\lambda \neq 0$ and $\psi \neq 0$, the OLS estimator for β_1 may be biased in another way. According to Card (1999), people with higher returns to education tend to acquire more schooling ($\psi > 0$), and hence

[2] We write $\overline{X \circ Y} = (1/n)\sum_{i=1}^n x_i y_i$, i.e. $\overline{X \circ Y}$ is the sample mean of the products $x_i y_i$.

[3] $\text{E}\,(\nu_{1i}|S_i) = 0$ implies that $\text{E}\,(\nu_{1i}S_i^2) = \text{E}\,[\text{E}\,(\nu_{1i}S_i^2|S_i)] = 0$.

[4] It can be shown that $\text{cov}(X, X^2) = \text{E}\,X^3 - \mu_X\sigma_X^2 - \mu_X^3$ and $\text{E}\,(X - \mu_X)^3 = \text{E}\,X^3 - 3\mu_X\text{E}\,X^2 + 2\mu_X^2 = \text{E}\,X^3 - \mu_X^3 - 3\mu_X\sigma_X^2$, as $\text{E}\,X^2 = \sigma_X^2 + \mu_X^2$. If X is symmetric, $\text{E}\,(X - \mu_X)^3 = 0$, and it follows that $\text{cov}(X, X^2) = 2\mu_X\sigma_X^2$.

a cross-sectional regression of earnings on schooling yields an upward-biased estimate of the average marginal return to schooling β_1. Verbeek (2000), on the other hand, argues that the individual–specific returns to schooling are potentially higher for individuals with low levels of schooling ($\psi < 0$), and hence a downward bias in the OLS estimate for β_1 is to be expected. For now, we do not favor one interpretation over the other, but emphasize that in both situations the OLS estimator is potentially biased, which may be either upward or downward[5].

5.2.4 Optimizing behavior bias

Finally, a fourth source of possible bias of the OLS estimator in the schooling equation is due to the optimizing behavior of individuals. This is discussed to some extent in Griliches (1977) and Card (1999, 2001). Schooling can be regarded as the result of optimizing behavior of individuals or households. Individuals try to reach an optimal schooling decision by maximizing 'wealth' or 'utility' based on anticipated earnings, that depends on schooling, ability, other unknown factors, and certain (opportunity) costs of schooling (depending on for instance interest rates and tuition fees). Garen (1984) views this as a self-selection problem with a continuous choice variable. Uusitalo (1999) and Harmon and Walker (1995) argue that this behavior could induce a positive bias in the OLS estimator. Griliches (1977) shows that it might as well lead to a downward bias. A more extensive discussion is beyond the scope of this manuscript and we refer to the aforementioned works.

5.3 IV estimation of the returns to education

Given the divergent and a priori unknown sources of potential regressor-error dependencies in estimating the return to education, it is not an easy task to find appropriate instruments that alleviate regressor-error dependencies in model

[5]Card (1999, 2001) shows for model (5.2) that presence of measurement error in S_i biases the OLS estimate for β_1 towards zero, implying that a relatively small amount of measurement error may (partly) offset a modest upward ability bias, depending on the magnitudes of λ and $\psi\mu_S$. This can also be seen from the example in subsection 5.2.2

(5.1). Card (1999, 2001) gives an overview of recent studies that use instrumental variables to estimate the return to schooling. He distinguishes two sets of instrumental variables that are commonly used: (1) those that are based on institutional features of the school system and (2) those that are based on family background characteristics, both of which are discussed next.

5.3.1 Institutional features of the schooling system

When instrumental variables based on institutional features are used, the resulting IV estimates are approximately 30% higher than the corresponding OLS results. This finding does not agree with current beliefs in the literature about the traditional ability bias. Card (1999, 2001) provides four explanations. Firstly, instruments based on institutional features of the schooling system may not be truly exogenous, since a direct effect of the instruments on earnings may exist. For instance, 'college proximity' is sometimes used as an instrument (see e.g. section 5.4), but it may have a direct effect on earnings, since families that place a strong emphasis on education may choose to live near a college, while their children may have higher abilities and/or motivation to achieve labor market success (cf. Verbeek, 2000). Bound and Jaeger (1996) argue that the quarter-of-birth dummy instruments used in Angrist and Krueger's (1991) study may have a direct association with the dependent variable, as their results suggest that the association between quarter-of-birth and earnings is too strong to be fully explained by school attendance laws. Card (1999)[6] argues that instruments based on schooling reforms (treatment), such as changes in compulsory school attendance laws, are biased further upward compared to OLS because of unobserved differences between the characteristics of the treated and non-treated group, since these reform treatments are often not random. Bound, Jaeger and Baker (1995) argue that Angrist and Krueger use a large number of weak instruments and show that, in finite samples, IV estimates based on weak instruments are biased in the same direction as OLS.

[6] par. 3.4 (p.1819-p.1822), and p.1841.

Secondly, the downward bias in OLS can be a result of error in the measurement of education. However, the strength of this effect is doubtful in view of Card (1999) who argues that it is unlikely that measurement error alone can account for the large positive gap between the IV and OLS estimates.

Thirdly, factors like compulsory schooling or schooling availability are most likely to affect individuals who otherwise would have had relatively low schooling. If, because of potential heterogeneity, these individuals have higher than average marginal returns to schooling, then instruments based on these variables tend to recover the returns to education for a subset of individuals with relatively high returns to education, resulting in estimates higher than OLS. Uusitalo (1999) notes in this respect that presence of heterogeneity in the coefficient of the returns to education yields an additional error term $\nu_{1i} S_i$. Since the instrument Z_i is correlated with S_i, it cannot be uncorrelated with the error term of the wage equation.

5.3.2 Family background

The second type of instrumental variables commonly used are instruments based on family background characteristics, for instance measures on education levels of family members. The use of these variables as instruments is motivated by the fact that children's education tend to exhibit a high correlation with parents' education. However, Card (1999) shows that when the OLS estimator is biased upward because of unobserved ability, the bias in the IV estimator is at least as large, and potentially larger, depending on the strength of the instruments and its possible direct effect on the dependent variable. Hence, if the OLS estimator is biased upward, one would expect that an IV estimator based on family background is biased upward even more. For a more detailed discussion we refer to Card (1999).

5.3.3 Alternative, non-IV approaches

Although using instrumental variables is a common way to solve for omitted 'ability' bias, some other methods are available (see e.g. Harmon and

Walker, 1995). If good measures for 'ability' are available, they can be included as proxy variables in the wage equation (5.1). Then, if no other sources of regressor-error dependencies or measurement error problems remain, standard OLS techniques can be applied to estimate the return-to-schooling effect. This approach, however, is doubtful. Empirical results using this approach suggest an upward 'ability' bias in least-squares estimates (see Blackburn and Neumark, 1993, Wooldridge, 2002). Griliches (1977), however, argues that it is difficult to find a good measure for ability. Often proxies for unobserved ability are measures of IQ. If, however, unobserved ability has no relation with IQ, but is instead related to, say, 'motivation', any proxy for ability based on IQ induces large measurement error biases in the OLS estimator, while unobserved ability may still not be accounted for (see also the discussion in section 5.2.2).

A particular powerful approach to address regressor-error dependencies in schooling models is to use data on twins (or siblings) (Card, 1999). This approach attempts to eliminate possible omitted variable biases by assuming that some of the unobserved factors (e.g. ability or motivation) are identical within families (or twin/sibling pairs). In this case, differences of levels of schooling and education for the twins or siblings can be exploited to estimate the effect of education on wage. Card (1999) gives an overview of several studies that use twin-data. He concludes that under the assumption that identical twins have identical abilities, the within-family estimator gives a consistent estimate for the average marginal returns to schooling. Furthermore, this estimator can be corrected for measurement error. Card (1999) concludes from his survey that the OLS estimator obeys a slight upward-bias of the order of $10\% - 15\%$. A drawback of these methods is the (possible) lack of generalization to non-twins and the potential failure of the identical abilities assumptions for identical twins and siblings. If the assumption does not hold, twin studies might still overestimate to some extent the effect of education on earnings. In a recent study, Hertz (2003) also finds that the OLS results are biased upward. His results are based on various measurement–error corrected, within–family estimators for South–African households.

The review of the literature on estimating the return to education in this and the previous sections demonstrates that IV estimation has produced a less than satisfactory solution to the endogeneity problem of the schooling effect. In the next section we present the LIV estimates for model (5.1) for three empirical applications and compare these estimates with classical IV (2SLS). We show that the LIV results are more stable across the three datasets and are more in line with recent evidence from twin studies. In addition, the LIV model fits the data fairly well, based on the diagnostics in section 4.5. We argue that the LIV estimates are to be preferred to the classical IV results in these applications.

5.4 Empirical results

In this section we present the results of three applications to examine the effect of education on income. Each of these three applications are based on previously published data. First we briefly describe the three datasets, where a more detailed description can be found in appendix 5A. We then estimate model (5.1) with latent instrumental variables and compare these results with the traditional IV and OLS estimates. Furthermore, we investigate the available observed instruments thoroughly, and conclude that the LIV results are to be preferred over IV and OLS.

5.4.1 Data description

NLSY data

The first dataset is a sample of 3010 men taken from the US National Longitudinal Survey of Young Men (NLSY) from 1976. This dataset is analyzed in Card (1995) and Verbeek (2000). The dataset contains several exogenous variables and one dummy instrumental variable measuring the presence of a nearby college, i.e. an instrument based institutional features of the school system.

Brabant data

The second dataset was originally sampled in 1952 from the Dutch province 'Noord-Brabant'. Thirty years later the same individuals were contacted to collect data on, among other things, educational level, income, and social background statistics. The labor market information used here is from 1983, and the dataset used contains observations on 833 men who had reached a stable labor market position. As with the NLSY dataset, several exogenous explanatory variables are available. We have two instrumental variables: measures on the educational level of the respondents' father and mother, i.e. family background characteristics (see also Hartog, 1988, for a more detailed description of the data).

PSID data

The third dataset contains data on 424 working, married white women between the ages 30 and 60 in 1975, and comes from the University of Michigan Panel Study of Income Dynamics (PSID), analyzed in Wooldridge (2002) and Mroz (1987). The labor market information is from 1975. This dataset has several exogenous variables. The available instruments are family background variables: the respondents' fathers and mothers level of education and the husbands level of education. For more details on the datasets and the used regressors and instruments, we refer to appendix 5A.

The three datasets differ on various key aspects (sample sizes, region, sex of respondents, year of labor market information), which makes direct comparison of the estimated regression coefficients superfluous. However, we compute the relative bias in OLS with respect to the LIV and IV estimates, which, as will become clear, allows for a straightforward comparison of the results across the three applications. The application of LIV with its assumption of discrete levels of the latent variable may well correspond to the existence of discrete levels of schooling, underlying the measured education variables, that are free of measurement error and that represent the levels of education that on would obtain regardless of ability, but is not predicated on that. Alternatively, as one reviewer to this study pointed out, LIV can be interpreted to identify 'latent twins' and using an analogue of the twin estimator, conceptually.

5.4.2 LIV results for schooling

We analyze all datasets with the LIV model and methods presented in chapters 3 and 4. We estimate the LIV model for $m = 2, ..., 5$ and with the inclusion of extra exogenous variables. We emphasize that the LIV model does *not* require the availability of instrumental variables, and the results in this subsection are obtained *without* using the available observed instruments mentioned in the previous subsection. In subsection 5.4.3 these are included in the model in order to examine their validity, using the methods in section 4.3. We also present here the results for the standard OLS estimator, the IV estimator, and LIV model fit diagnostics, but postpone a detailed discussion of the IV results until subsection 5.4.3.

Estimated coefficients

Table 5.1: Results of OLS, IV and LIV for the schooling coefficient for the three datasets. LIVx means that the LIV model is estimated with $m = x$ categories.

β_S	OLS	IV	LIV2	LIV3	LIV4	LIV5
NLSY	0.074	0.133	0.050	0.065	0.068	0.069
	(0.0035)	(0.0518)	(0.0099)	(0.0041)	(0.0040)	(0.0040)
Brabant	0.043	0.056	0.040	0.042	0.040	–
	(0.0044)	(0.0075)	(0.0051)	(0.0049)	(0.0049)	
PSID	0.102	0.073	0.134	0.099	0.099	0.096
	(0.0139)	(0.0321)	(0.0282)	(0.0160)	(0.0153)	(0.0142)

In table 5.1 we present the results for the estimated schooling coefficients for the datasets using OLS, IV, and LIV. It can be seen that for all specifications for m in the LIV model (denoted by LIVx), the resulting estimate for the schooling coefficient is below the OLS estimate, indicating a small upward bias in the

OLS estimate. On the other hand, the direction of the bias for the IV results using the observed instruments is not unanimous, and we discuss this in more detail in subsection 5.4.3. The only downward bias found by LIV is for the PSID data when $m = 2$. This can be expected if a dummy variable exists which is identical or nearly identical to the unobserved instrument. In this case, there is a situation of (almost) perfect multicollinearity in the second stage of the LIV model and the parameters are only nearly identified. This also explains why the results for $m = 2$ have larger standard deviations than what could have been expected and why relative large improvements in model fit occur for $m > 2$. In these applications several dummy regressors are present (see appendix 5A). In addition, the PSID dataset is the smallest we have and the likelihood surface may be less smooth in this case. For the Brabant data the maximized value of the likelihood is degenerate at $m = 5$ and no estimate for LIV5 is given in table 5.1. We found for $m > 5$ also degenerate solutions for the PSID data. Here the LIV method indicates that the number of instruments (number of categories) should not be too large. Overall it can be seen that the LIV results are fairly stable for different choices of m and we consider choosing the 'best' m next.

Choosing the number of categories of the latent instrument

As argued in the previous chapter, we choose among the different values for m by looking at the ICL criterion, and for comparison and validity we also present AIC3 and BIC in table 5.2. For the NLSY data the ICL statistic yields a minimum at $m = 4$ and AIC3 and BIC at $m = 5$. For the Brabant dataset ICL yields $m = 2$ and AIC3/BIC $m = 4$. All three measures give $m = 5$ for the PSID data. In accordance with recent evidence on the performance of the 'classical' selection criteria, we also find in two of the three cases that AIC-based statistics point to a larger number of categories for the discrete instrument. Importantly, it can be seen from table 5.1 and from results in appendix 5B that the estimated regression coefficients and the estimates for the schooling equation are not very different for the optimal values of m as indicated by ICL and AIC3/BIC. As we will show this result also holds for testing for (absence of) endogeneity. In the following we will only consider

Table 5.2: Computed values for BIC, AIC3 and ICL. Boldface values indicate the minimum value (row-wise), and a * denotes a degenerate solution.

		$m = 2$	$m = 3$	$m = 4$	$m = 5$
NLSY	BIC	5832.06	5404.04	5309.55	**5291.73**
	AIC3	5751.91	5313.86	5209.36	**5181.52**
	ICL	6942.75	5703.59	**5611.37**	5995.09
Brabant	BIC	1867.02	1837.67	**1835.73**	1849.18*
	AIC3	1799.967	1763.174	**1753.774**	1759.774*
	ICL	**1931.07**	1974.67	1990.93	2004.44*
PSID	BIC	1164.49	1023.99	1005.42	**905.26**
	AIC3	1103.498	956.894	932.227	**825.97**
	ICL	1199.23	1042.04	1020.93	**914.55**

the LIV results for the optimal number of categories for the latent instrument, i.e. $m = 4$ for NLSY, $m = 2$ for Brabant, and $m = 5$ for PSID.

Testing for endogeneity

Table 5.3 shows the results for the relative bias[7] in the estimated regression coefficient for schooling with respect to OLS for the IV and optimal LIV results. Furthermore, the test results for testing for absence of endogeneity are presented. We present the results for IV (2SLS) as well, but discuss the IV estimates and the used instruments in more detail later on. The test-statistics for LIV are computed without using the observed instrumental variables. The Hausman-test is based on comparing the complete vectors $\hat{\beta}_{\text{OLS}}$ and $\hat{\beta}_{\text{LIV}}$ (and $\hat{\beta}_{\text{IV}}$) (see appendix 5B for the estimates of the complete vector of regression coefficients). The Wald-test examines the covariance between the error terms of the main regression equation and the equation for the endogenous schooling (see section 4.6). This test cannot be computed for 2SLS.

Overall, we see that the differences between LIV and OLS are not large, which

[7]We computed this percentage as $100 \times (1 - \hat{\beta}_1^{\text{LIV}} / \hat{\beta}_1^{\text{OLS}})$ and $100 \times (1 - \hat{\beta}_1^{\text{IV}} / \hat{\beta}_1^{\text{OLS}})$.

Table 5.3: Relative biases with respect to OLS and results for Hausman- and Wald-tests to test for endogeneity (based on Hessian).

Data	Estimator	%Δ	H-test	W-test
NLSY	IV	-79.9	1.31	-
	Opt. LIV $m = 4$	7.9	9.28	9.31
	Opt. LIV $m = 5$	6.5	7.18	7.19
Brabant	IV	-30.1	4.35	-
	Opt. LIV $m = 2$	7.0	0.97	0.97
	Opt. LIV $m = 4$	7.0	2.63	2.63
PSID	IV	27.8	0.95	-
	Opt. LIV $m = 5$	5.5	4.20	4.53

is also indicated by the Hausman- and Wald tests (presented in the last two columns of table 5.3)[8]. Both tests give similar conclusions. The optimal LIV solutions for the NLSY data and the PSID data indicate a significant upward bias in OLS, but for the Brabant data the estimated value for β_1 by LIV (for $m = 2$) is not significantly different from OLS. Here, the classical IV estimator indicates a significant downward bias in OLS.

Before discussion the classical IV results in more detail, we first examine various diagnostics for the above presented LIV estimates, where we only report the results for the LIV model indicated by the (preferred) ICL-criteria and report, in case of the NLSY data and the Brabant data, only the results for the model selected by AIC3 when these are substantially different. We note that for the Brabant data the R^2 measure for the strength of the LIV instrument, as discussed in subsection 4.5.2, is substantially better for $m = 4$ than for $m = 2$, which is discussed in more detail in subsection 5.4.3.

[8]The critical 5%-value of a χ^2_1 distribution is 3.84.

Diagnostics: outliers, influential observations, normality and heteroscedasticity

We examined the various diagnostics presented in the previous section to investigate the fit of the (optimal) LIV model and to identify potential outliers and influential observations. Residual plots for the three datasets are given in figure 5.1.

For the NLSY data ($n = 3010$), residual checks did not reveal heteroscedasticity, and residuals had a skewness of -0.28 and a kurtosis of 3.5. All standardized residuals were smaller than (in absolute value) 4.5. Examining the outliers and influential observations diagnostics did not identify highly unusual data.

For the PSID data ($n = 424$) there is evidence of weak heteroscedasticity for the variable ‘experience’, but this effect is rather small. The residuals are slightly skewed (-0.26) and are leptokurtic[9] (kurtosis is 5.1). One observation was identified as an influential observation, but no outliers are present. When this observation is removed results and conclusions do not change, and all standardized residuals are smaller (in absolute value) than 4.

As for the PSID data, the results for the Brabant data ($n = 833$) indicate slight evidence of weak heteroscedasticity, here for the dummy variable whether the father is self employed at the age of 12. For this dataset, the residuals are more skewed (skewness is -1.25) and more leptokurtic (kurtosis is 12.7). However, examination for potential outliers and influential data identifies three observations that clearly do not ‘fit’ the rest of the data. We re-estimated the model without these three observations, and found that the estimates and test statistics are not strongly affected. The Hausman- and Wald statistic for the $m = 4$ solution, however, now become 3.47 and 3.48, respectively, which are both significant at $\alpha = 0.10$. After omission of these outlying data, the residuals are less skewed and leptokurtic. All but four of the absolute values of the standardized residuals are smaller than 4.5, with a maximum of 5.9. We note that

[9] A distribution with a high peak is called leptokurtic (Weisstein, 2004b).

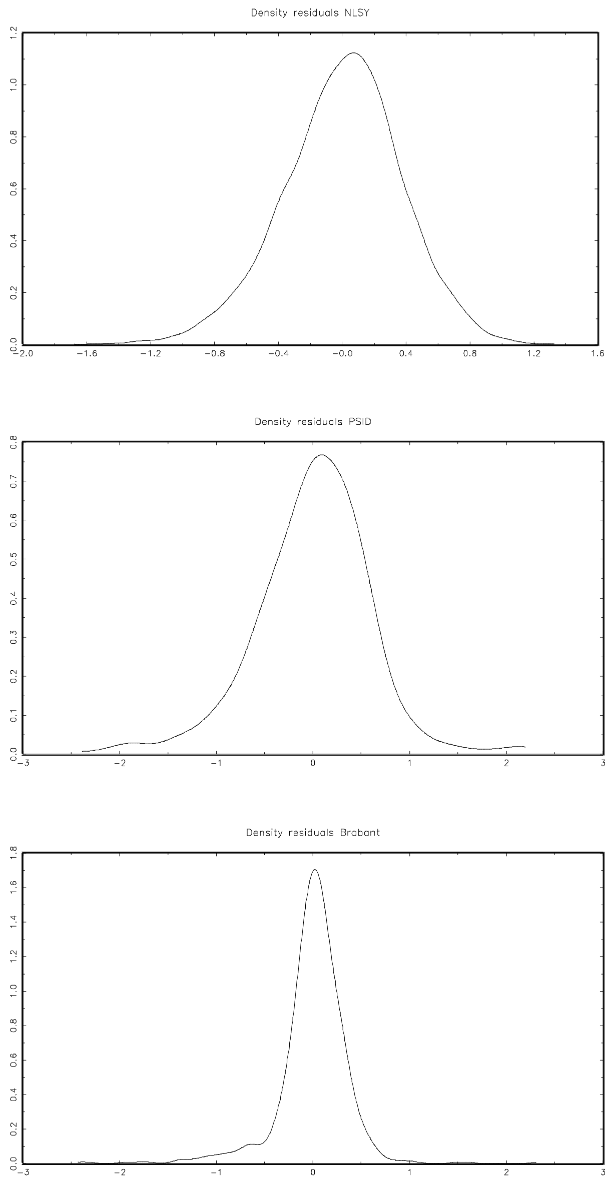

Figure 5.1: Residuals.

the measures for skewness and kurtosis are based on higher order moments, which are known to be sensitive to outliers.

5.4.3 Relative biases and comparison with classical IV

From the relative percentage bias in OLS with respect to the optimal LIV and IV estimates in table 5.3 (the column indicated by $\%\Delta$), it can be seen that the LIV method reveals an upward bias of OLS ranging from $5.5\% - 8\%$. When traditional IV is used, the conclusions are very different for the three studies, ranging from an $\approx$ 80% downward bias to an $\approx$ 30% upward bias in OLS. For the NLSY data, the IV estimate for the return to schooling, based on a dummy for college proximity, is about 80% higher than OLS and equal to (approximately) 0.13 (0.052). For the Brabant data, we find that the IV estimate is 0.056 (0.008), which is also substantially higher ($\approx$ 30%) than OLS. Here the instruments are the levels of education of the respondents' parents. Using similar instrumental variables, we find for the PSID data an *up*ward bias of $\approx$ 30% in the OLS estimate. It can be seen that in all cases, the IV estimate has a standard deviation that is substantially higher than OLS. The instability of the 2SLS results and the high standard deviations may be a result of weak and/or endogenous instruments, which is investigated in the next subsection.

For the Brabant data and the PSID data we had more instruments available than necessary for identification, allowing us to examine the sensitivity of IV to different choices for the set of instruments[10]. When only mothers' education is used as IV in the Brabant data, i.e. the number of instruments is decreased, the estimated coefficient for the return to education becomes 0.059 (0.009), which is slightly higher than the estimate obtained from using the full set of instruments. Similarly, in the PSID data we also have husband's education as an additional covariate. When this variable is included in the set of instruments, i.e., the number of instruments is increased, the IV estimate becomes 0.065 (0.023). This value is about 11% lower than the IV results for the smaller set of instruments. These results indicate that, in particular for the PSID data, the

[10]This can be seen as changing the number of categories of the latent instrument in the LIV model.

2SLS estimates may be sensitive to different choices for the set of instruments. In the following we examine the strength of the available observed instruments.

Strength of the available observed instruments

Table 5.4: Results strength of observed versus predicted LIV instruments. Instruments NLSY: 'Nearc', Brabant: 'FatherEd' and 'MotherEd', respectively, and PSID: 'husbanded', 'mothered', and 'fathered', respectively (based on Hessian).

Data	Method	ΔR^2	$\gamma_{2z_{21}}$	$\gamma_{2z_{22}}$	$\gamma_{2z_{23}}$	Test
NLSY	Obs IV	0.0029	0.34			–
			(0.11)			
	LIV4 IV	0.7503	0.15			6.83
			(0.06)			
	LIV5 IV	0.7976	0.16			7.85
			(0.06)			
Brabant	Obs IV	0.0922	1.08	1.27		–
			(0.18)	(0.22)		
	LIV2 IV	0.3906	0.89	1.02		104.41
			(0.16)	(0.19)		
	LIV4 IV	0.5247	0.61	0.93		48.48
			(0.14)	(0.21)		
PSID	Obs IV	0.3225	0.34	0.12	0.10	–
			(0.03)	(0.03)	(0.03)	
	LIV5 IV	0.8312	0.04	-0.01	0.02	14.87
			(0.01)	(0.01)	(0.01)	

Table 5.4 shows the results of the diagnostics proposed in chapter 4 that examine the strength of the observed instruments. Investigating the strength of the observed instruments is important when using classical IV (2SLS) estimation. We examine the R^2 of the first-stage regression for the observed and optimal LIV instruments (see subsection 4.5.2), and discuss the results of the test proposed in subsection 4.3.1. Here we used all available observed instruments[11]

[11] In fact, we estimate model (4.12). These estimation results are also used in the next subsection to examine the exogeneity of the observed instruments. Including the observed instruments

and present conclusions only for the optimal *m* given in table 5.2.

The third column of table 5.4 reports the difference in R^2 of the regression of schooling on the explanatory variables and the available observed instruments, or, in case of LIV, the optimal LIV instruments, and the R^2 of the regression of schooling on the exogenous explanatory variables *only*. Hence, a large increase in R^2 indicates that the instruments explain a substantial amount of the variance in the endogenous schooling variable. It can be seen that in particular for the NLSY data the observed instrument 'Nearc' appears to be weak. The family background instruments (Brabant and PSID data) explain a larger part of the variance in schooling, in particular for the PSID dataset. However, the increase in R^2 is in all cases substantially larger when using the optimal LIV instruments. It follows that the optimal LIV instruments do a much better job in explaining the variance of x than the available observed instruments. These findings explain the loss of efficiency in the 2SLS estimates for the regression coefficients in table 5.1, where the IV estimated standard deviations are (0.052), (0.008), and (0.032) and, respectively, 14.8, 1.7, and 2.3 times higher than the OLS standard deviations. Not surprisingly, the estimated standard deviations for (the optimal) LIV estimates are only 1.14, 1.16, and 1.02 times the OLS estimated standard deviations.

The results for the Wald-test to test for a zero-effect of the observed instruments on schooling is given in the last column of table 5.4, and the reported coefficients in the columns γ_{2z_2} are the estimated direct effects of the observed instruments on schooling. For instance, for the NLSY data it can be seen that individuals who lived near a college have slightly more education, and from the Brabant and PSID data it follows that parents education is positively related to the years of schooling of their children. Under the null hypothesis $H_0 : \gamma_{2z_2} = 0$, the test-statistic has approximately a χ^2-distribution with degrees of freedom 1, 2, and 3, for, respectively, the NLSY, Brabant, and PSID data. It can be seen that in all cases the null hypothesis is rejected, indicat-

yields the following relative biases (%Δ): for the NLSY data 7.9% ($m = 4$) and 6.6 % ($m = 5$), for the Brabant data 6.6% ($m = 2$) and 9.4% ($m = 4$), and for the PSID data 9.0% ($m = 5$).

ing that the observed instruments have a non-zero direct effect on schooling[12]. For the NLSY data, however, the P-values are 0.009 ($m = 4$) and 0.005 ($m = 5$), which provides evidence that the instrument used is considerable weak, given the remarks in subsection 4.3.1 and the substantial sample size of this dataset. For the Brabant and PSID data we also tested H_0 for each instrument separately. Only for the instrument 'mothered' in the PSID data the null hypothesis of zero effect on schooling was not rejected. Although in all cases the null hypothesis of a (joint) zero effect is rejected using $\alpha = 0.01$, the incremental R^2's were not large, in particular for the NLSY data. Here the available instrument seems to be weak. However, in all cases the optimal LIV instruments were found to be substantially stronger than the available observed instruments. Their exogeneity is considered next.

Examining exogeneity of available observed instruments

Table 5.5: Results endogeneity test of available observed instruments. Instruments NLSY: 'Nearc', Brabant: 'FatherEd' and 'MotherEd', respectively, and PSID: 'husbanded', 'mothered', and 'fathered', respectively (based on Hessian).

Data	Method	$\beta_{2z_{21}}$	$\beta_{2z_{22}}$	$\beta_{2z_{23}}$	Test
NLSY	Opt. LIV $m = 4$	0.022			1.810
		(0.02)			
	Opt. LIV $m = 5$	0.021			1.760
		(0.02)			
Brabant	Opt. LIV $m = 2$	0.015	0.055		6.075
		(0.02)	(0.03)		
	Opt. LIV $m = 4$	0.017	0.058		6.920
		(0.02)	(0.03)		
PSID	Opt. LIV $m = 5$	-0.019	-0.006	0.000	2.731
		(0.01)	(0.01)	(0.01)	

In table 5.5 we present the results of the Wald-test for testing whether the coefficient of the direct effect of the observed instruments on the dependent vari-

[12]The 2SLS model is not identified under the null hypothesis.

able (wage) is zero. A nonzero effect would violate the exogeneity assumption of the instrument. We emphasize that the test in the previous subsection (table 5.4) tested whether the effect of the instruments on the endogenous regressor was zero.

The estimated coefficients are reported in the columns indicated by β_{2z_2} and the values of the test-statistic are given in the last column. The estimated coefficients β_{2z_2} are the direct effects of the instruments on the dependent variable wage. The degrees of freedom of the χ^2 null-distribution are, as before, 1, 2, and 3, for, respectively, the NLSY, Brabant, and PSID data. For the Brabant data the null hypothesis of no (joint) direct effect of the observed instruments on the dependent variable is rejected for $\alpha = 0.05$, suggesting that these instrumental variables are not exogenous, i.e. parental education levels have a significant positive effect on the respondent's wage. Performance of the 2SLS estimator critically relies on the exogeneity of the used instruments and these results suggest that the 2SLS estimates for the Brabant data are to be distrusted. For the other two datasets there was no evidence of significant non-zero effects of the observed instruments on the dependent variable.

5.4.4 Wrap-up

Our results illustrate the difficulties associated with IV estimation in these applications. The conclusions for the three datasets with respect to the magnitude and sign of the bias in the estimated OLS coefficient for schooling differ highly, even with a similar set of instruments. The results on the validity of the available observed instruments in the previous subsections may explain part of this variability.

First of all, the instrument 'Nearc' for the NLSY data was found to be the weakest available instrumental variable, inducing only a minor increase in the R^2 of the first stage regression. The simulation results in subsection 4.4.1 showed that presence of weak instruments results in large swings in the 2SLS estimates[13], which may account for the large downward bias in the OLS estimate found for the NLSY data (see also Card, 1999, 2001). Accordingly, the 2SLS estimator suffers from large standard deviations in these cases. Secondly, the family background instruments used for the Brabant data were found to have a direct effect on the dependent variable, in which case the bias in the 2SLS estimates may be of the same sign and potentially larger in magnitude than the bias in OLS, which explains the downward bias in OLS found in the Brabant application when using 2SLS (see also Bound, Jaeger, and Baker, 1995). Finally, our results suggest that the instruments for the PSID data are the 'best' among the three applications, although evidence suggests that they are somewhat weak. The 2SLS results for the PSID data do indicate an upward bias in OLS, but the magnitude is much larger than for the LIV model.

Card (1999) argues that instruments based on family background characteristics are likely to be endogenous, in the presence of omitted ability, which was supported by our findings for the Brabant data. We did not find evidence for this hypothesis for the PSID data, where the respondent's husband, father, and mother's levels of education are the available observed instruments. One explanation is that the power of the test is lower for the PSID data because of the smaller sample size. Furthermore, the PSID data contains labor market information on women obtained in 1976, while in that period education, income, and other labor market issues, may have been less of an issue for women than for their male siblings. Hence, it can be expected that family education has a lower correlation with the respondents data.

These results are contrary to the optimal LIV solutions, which are more efficient than standard IV since the LIV method optimally estimates instruments from the available data. Furthermore, the best available evidence from the latest studies on identical twins suggests a small upward bias on the order of 10–15% in the OLS estimator (cf. Card, 1999), which is not supported by the standard IV estimates from the three datasets analyzed here. Our estimates

[13]In fact, the results in table 4.1 for the weak instrument cases report median bias and IQR instead of mean bias and standard deviation because of the presence of too many outliers.

have the same order of magnitude found in the twin studies but do not fully recover the 10% difference. A reason for this result might be that estimating the model by simple OLS yields in general only a modest fit (the OLS results presented in table 5.1 have R-squares of respectively 29%, 23%, and 17%), i.e. the regressors do not explain a large part of the variance in wage. The fact that LIV finds a smaller positive bias might also indicate that a part of the positive ability bias is offset by negative biases due to e.g. measurement error or heterogeneity, which is expected to be less in the twin studies[14]. Further, in the twin studies there may still be a limited amount of unobserved ability if the abilities of twins and siblings differ.

5.5 Conclusions

The studies of Card (1999, 2001) clearly indicate the difficulties associated with applying standard IV methods to estimate the returns to education. The results are often found to be counterintuitive and different across studies. Furthermore, in many instances it can be questioned whether the instruments used were 'valid'. Unfortunately, classical instrumental variable methods do not allow for straightforward testing of the validity of a specific instrument. We show that the LIV method can be successfully applied to solve these problems. The OLS estimates are found to be biased upward by about 7%. Equally important, the available observed instruments that have been used seem to be mostly inadequate, and produce results that are both more biased than the OLS results and have much lower efficiency, see table 5.6.

The advantage of the LIV approach is that no observable instruments are needed. Furthermore, once estimates have been obtained, endogeneity can be tested for in a straightforward manner. We showed that for the different specifications of m and across three different datasets the estimates are consistent. For the

[14]For instance, let the true schooling effect be $\beta_S = 10$. With a $+4$ ability bias and a -2 measurement error bias, the estimated schooling effect by OLS is 12, resulting in an upward bias of $\approx$ 16% in OLS. With fewer measurement error, e.g. -1, the OLS estimate is 13, which yields an upward bias of $\approx$ 23%.

Table 5.6: Summary main conclusions LIV results effect of education on earnings.

	Data		
	NLSY	Brabant	PSID
Sample size	3010	833	424
Rel. bias OLS by LIV	7.9%	7%	5.5 %
Rel. bias OLS by IV	-79.9%	-30.1%	27.8%
Test for endogeneity	+	-	+
Instrument strength	weak	moderate	moderate
Instrument exogeneity	exogenous	endogenous	exogenous

NSLY and the PSID dataset we find significant evidence of an 'ability' bias[15]. Furthermore the standard errors of the estimates are much smaller than the standard errors for standard IV, and not much larger than OLS. Because of the relative large number of observations in the NLSY data, it is to be expected that the power of the Hausman- and Wald-test is larger. In using LIV to test for endogeneity it is recommended to use datasets of substantial size to ensure a reasonable power. We proposed several diagnostics that may complete an analysis using LIV. We do not find any evidence that the LIV models used for the three applications here do not fit the data well. Especially, in view of the large samples sizes for the three datasets, the small deviations from the assumptions that were found may not pose a problem in making inferences.

The relative size and magnitude of the bias in the OLS estimator that was found is somewhat smaller, but still close to the numbers reported in Card (1999) for the twin studies: 6–8% for all three datasets. This shows considerable convergent validity and it lends additional credibility to the LIV approach. In the next chapter we consider estimation of multilevel models in presence of endogenous regressors.

[15] After deleting three outliers, the tests used also point out a significant ($\alpha = 0.10$) upward bias in OLS for the Brabant data (subsection 5.4.2).

Appendix 5A Descriptive statistics datasets used

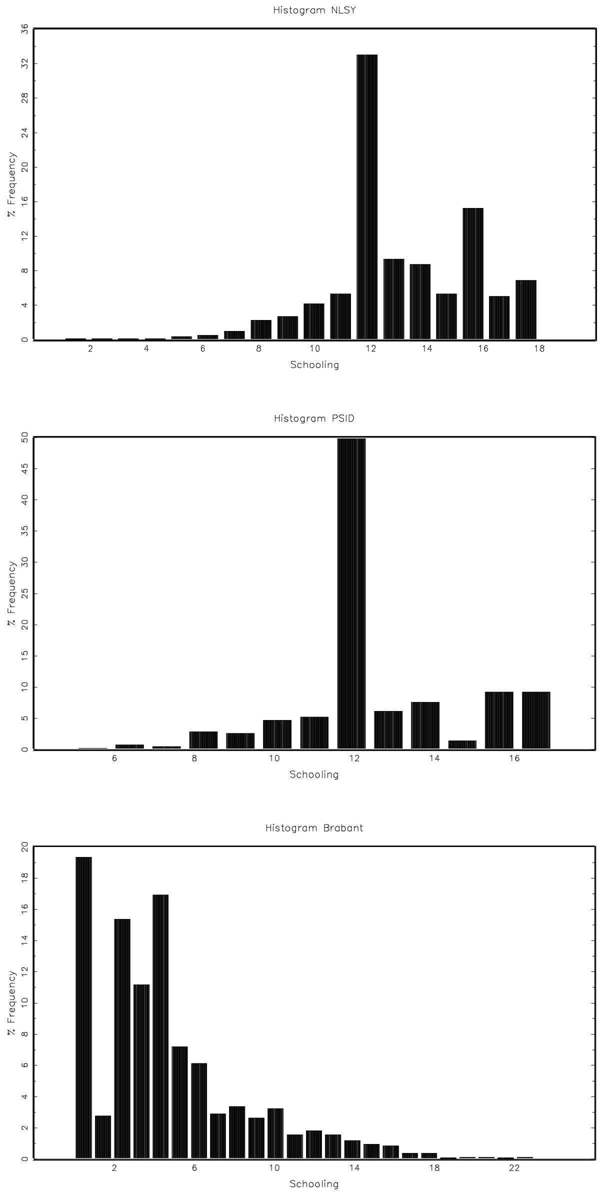

Figure 5A.1: Histogram of 'Schooling'.

5A.1 NLSY data

The total sample consists of 3010 men taken from the National Longitudinal Survey of Young Men (see Verbeek, 2000, and Card, 1995)[16]. In this survey, a group of individuals in the age of $14-24$ years is followed since 1966. The labour market information used is from 1976. In this year, the individuals had on average a little more than 13 years of schooling, with a maximum of 18 years (see figure 5A.1). The average working experience was about 8.86 years (in 1976 those men aged $24-34$) with an average hourly wage rate of \$5.77. The variables used can be found in table 5A.1. We used the values centered around the mean for schooling in estimation.

5A.2 Brabant data

The initial dataset[17] used in this paper consisted of 839 observations, but we deleted 5 observations with very low wages (log hourly wages < 0). Another observation with an extreme large reported wage was also removed ($> 9\times$ IQR from median). This data was collected in 1983 in the Netherlands' southern province of Noord-Brabant. At that time the average age of the men in the sample was about 43. This cohort was confronted with compulsonary schooling until 12 years of age. The schooling measure used is the number of post–compulsonary years of schooling; on average 4.35 years (see figure 5A.1). The average hourly wage was Dfl. 16.72 and the individuals had, on average, 25 years of work experience at the time of the survey. See table 5A.2 for more information on the variables used. As before, we centered the schooling variable around the mean.

5A.3 PSID data

As for the Brabant dataset, we removed a few observations prior to data analysis: four observations had an obvious lower (log) wage ($<< -1$) than the rest and were not used for estimation. This data come from the University of Michigan Panel Study of Income Dynamics (PSID)[18], obtained in 1976 (also used in Mroz, 1987). The sample consists of working married white women, who were aged in between 30 and 60 in 1975. They earned on average \$4.18 per hour. The women reported an average 12.7 years of schooling (see figure 5A.1) and a little over 13 years of labor market experience. For a detailed description of the used variables, see table 5A.3, where for estimation, the schooling variable was mean-centered.

[16] http://www.econ.kuleuven.ac.be/GME/.

[17] We thank Hans van Ophem (University of Amsterdam) for making this dataset available to us.

[18] http://mitpress.mit.edu/Wooldridge-EconAnalysis.

Table 5A.1: Descriptive statistics NLSY dataset ($n = 3010$).

Variable	Description	Mean	Std.
Regressors			
constant (β_0)	Model constant	-	-
schooling (β_1)	Years of schooling in 1976	13.26	2.68
experience (β_2)	Potential experience	8.86	4.14
black (β_3)	Equals 1 if black	0.23	0.42
smsa (β_4)	Equals 1 if lived in metropolitan area in 1976	0.71	0.45
south (β_5)	Equals 1 if lived in south in 1976	0.40	0.49
Dependent			
log wage	Logarithm of hourly wage	6.26	0.44
Instruments			
Nearc	Grew up near a 4 year college	0.68	0.47

Table 5A.2: Descriptive statistics Brabant dataset ($n = 833$).

Variable	Description	Mean	Std.
Regressors			
constant (β_0)	Model constant	-	-
schooling (β_1)	Years of schooling after age 12	4.35	4.00
experience (β_2)	Potential experience	25.52	4.19
nr. children (β_3)	Number of children present at age 12	4.91	2.68
av. mark (β_4)	Average school mark in final year of primary education	5.62	1.42
anti-social (β_5)	Equals 1 comes from antisocial background	0.10	0.29
fself (β_6)	Equals 1 if father is self employed at age 12	0.31	0.46
Dependent			
log wage	Logarithm of hourly wage	2.70	0.42
Instruments			
Father Ed.	Education level father	2.35	0.70
Mother Ed.	Education level mother	2.22	0.54
	(levels: $1 - 6$, higher categories = higher education)		

Table 5A.3: Descriptive statistics PSID dataset ($n = 424$).

Variable	Description	Mean	Std.
Regressors			
constant (β_0)	Model constant		
schooling (β_1)	Years of schooling	12.66	2.29
experience (β_2)	Actual labor market experience	13.09	8.05
kidslt6 (β_3)	Number of children younger than 6	0.14	0.39
kidsgr6 (β_4)	Number of children older than 6	1.34	1.32
unempl (β_5)	Unemployment rate in county of residence	8.54	3.04
city (β_6)	Equals 1 if lives in SMSA	0.64	0.48
nwincome (β_7)	Family income less total income wife / 1000	18.992	10.62
Dependent			
log wage	Logarithm of hourly wage	1.22	0.67
Instruments			
father ed.	Years of schooling father	8.80	3.57
mother ed.	Years of schooling mother	9.24	3.37
husband ed.	Years of schooling husband	12.50	3.02

Appendix 5B Results optimal LIV model for the three datasets

Table 5B.1: *NLSY data.* Results for OLS, IV, and optimal LIV. Here β_0 is the constant, β_1 is ‘schooling’, β_2 is ‘experience’, β_3 is ‘black’, β_4 is ‘smsa’, and β_5 is ‘south’.

	β_0	β_1	β_2	β_3	β_4	β_5	σ_ϵ^2
OLS	6.262	0.074	0.039	-0.188	0.165	-0.129	0.142
	(0.007)	(0.004)	(0.002)	(0.018)	(0.016)	(0.015)	
IV	6.262	0.133	0.040	-0.103	0.109	-0.100	0.164
	(0.007)	(0.052)	(0.002)	(0.077)	(0.051)	(0.030)	
Opt. LIV	6.262	0.068	0.039	-0.197	0.170	-0.132	0.142
$m = 4$	(0.007)	(0.004)	(0.002)	(0.018)	(0.016)	(0.015)	
Opt. LIV	6.262	0.069	0.039	-0.196	0.169	-0.132	0.142
$m = 5$	(0.007)	(0.004)	(0.002)	(0.018)	(0.016)	(0.015)	

Table 5B.2: *NLSY data.* Optimal LIV results for the estimated group probabilities π_j, group sizes λ_j (in *italics*), and γ_2. Here γ_{21} is 'age', γ_{22} is 'black', γ_{23} is 'smsa', and γ_{24} is 'south'.

π_1	π_2	π_3	π_4	π_5	γ_1	γ_2	γ_3	γ_4	σ_ν^2	$\rho_{\epsilon\nu}$
-8.50	-4.70	-1.03	2.96		0.01	-0.60	0.29	-0.14	1.10	0.091
(0.34)	(0.12)	(0.03)	(0.04)		(0.01)	(0.07)	(0.06)	(0.06)		
0.01	*0.07*	*0.59*	*0.34*							
-8.25	-4.53	-1.07	2.31	3.90	-0.01	-0.61	0.28	-0.09	0.84	0.091
(0.29)	(0.10)	(0.03)	(0.08)	(0.13)	(0.01)	(0.07)	(0.06)	(0.06)		
0.01	*0.07*	*0.56*	*0.24*	*0.12*						

Table 5B.3: *Brabant data.* Results for OLS, IV, and optimal LIV. Here β_0 is the constant, β_1 is 'schooling', β_2 is 'experience', β_3 is 'nr.children', β_4 is 'av. school mark', β_5 is 'anti-social', and β_6 is 'fself'.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	σ_ϵ^2
OLS	2.701	0.043	0.004	0.004	0.033	-0.141	0.008	0.133
	(0.013)	(0.004)	(0.004)	(0.005)	(0.010)	(0.029)	(0.045)	
IV	2.701	0.056	-0.004	0.005	0.011	-0.160	0.051	0.137
	(0.013)	(0.008)	(0.006)	(0.005)	(0.015)	(0.031)	(0.050)	
Opt. LIV	2.701	0.040	0.005	0.003	0.038	-0.137	-0.001	0.132
$m = 2$	(0.013)	(0.005)	(0.004)	(0.005)	(0.011)	(0.029)	(0.046)	
Opt. LIV	2.701	0.040	0.006	0.003	0.039	-0.136	-0.004	0.132
$m = 4$	(0.013)	(0.005)	(0.004)	(0.005)	(0.011)	(0.029)	(0.045)	

Table 5B.4: *Brabant data.* Optimal LIV results for the estimated group probabilities π_j, group sizes λ_j (in *italics*), and γ_2. Here γ_{21} is 'age', γ_{22} is 'nr. children', γ_{23} is 'av. school mark', γ_{24} is 'anti-social', and γ_{25} is 'fself'.

π_1	π_2	π_3	π_4	γ_1	γ_2	γ_3	γ_4	γ_5	σ_ν^2	$\rho_{\epsilon\nu}$
-1.00	6.38			0.32	-0.02	0.71	0.82	-1.66	4.39	0.057
(0.10)	(0.30)			(0.03)	(0.03)	(0.07)	(0.19)	(0.28)		
0.86	*0.14*									
-1.62	2.63	6.92	11.17	0.36	0.00	0.61	0.59	-1.48	2.35	0.105
(0.10)	(0.32)	(0.45)	(0.81)	(0.03)	(0.03)	(0.06)	(0.17)	(0.23)		
0.72	*0.19*	*0.08*	*0.01*							

Table 5B.5: *PSID data.* Results for OLS, IV, and optimal LIV. Here β_0 is the constant, β_1 is 'schooling', β_2 is 'experience', β_3 is 'kidslt6', β_4 is 'kidsgr6', β_5 is 'unempl', β_6 is 'city', and β_7 is 'nwincome'.

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	σ_ϵ^2
OLS	1.218	0.102	0.014	0.004	-0.007	-0.002	0.029	0.004	0.376
	(0.030)	(0.014)	(0.004)	(0.079)	(0.025)	(0.010)	(0.065)	(0.003)	
IV	1.218	0.073	0.014	0.030	-0.012	0.000	0.038	0.005	0.380
	(0.030)	(0.032)	(0.004)	(0.083)	(0.025)	(0.010)	(0.066)	(0.004)	
Opt. LIV	1.218	0.096	0.014	0.010	-0.008	-0.001	0.031	0.004	0.369
$m = 5$	(0.029)	(0.014)	(0.004)	(0.078)	(0.024)	(0.010)	(0.065)	(0.003)	

Table 5B.6: *PSID data.* Optimal LIV results for the estimated group probabilities π_j, group sizes λ_j (in *italics*), and γ_2. Here γ_{21} is 'experience', γ_{22} is 'kidslt6', γ_{23} is 'kidsgr6', γ_{24} is 'unempl', γ_{25} is 'city', and γ_{26} is 'nwincome'.

π_1	π_2	π_3	π_4	π_5	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	σ_ν^2	$\rho_{\epsilon\nu}$
-5.25	-2.97	-0.64	1.39	3.76	0.00	0.11	0.00	0.00	0.05	0.01	0.22	0.092
(0.11)	(0.09)	(0.03)	(0.10)	(0.06)	(0.00)	(0.07)	(0.02)	(0.01)	(0.06)	(0.00)		
0.04	*0.07*	*0.60*	*0.09*	*0.19*								